

畜禽基因组选择研究进展

张哲, 张勤*, 丁向东

中国农业大学动物科技学院, 农业部动物遗传育种与繁殖重点开放实验室, 畜禽育种国家工程实验室, 北京 100193

* 联系人, E-mail: qzhang@cau.edu.cn

2011-03-08 收稿, 2011-06-09 接受

引进国际先进农业科学技术计划(2011-G2A, 2010-C14)、国家自然科学基金(30800776)、转基因生物新品种培育重大专项(2009ZX08009-146B)和中国农业大学研究生科研创新专项资助

摘要 基因组选择是同时对高密度全基因组标记的选择, 也是一种标记辅助选择. 此方法能够在提高育种值估计准确性的同时, 通过早期选种缩短世代间隔、降低近交、加速遗传进展. 随着畜禽高密度全基因组 SNP 芯片的问世, 基因组选择方法正在影响着传统的畜禽遗传评估体系. 在多数奶业发达国家, 基因组选择方法正逐渐取代传统奶牛遗传评估方法, 成为奶牛遗传评估的“标准”方法. 近年来, 针对基因组选择的方法、影响因素、应用策略和育种方案的研究大量涌现. 基因组选择已经成为当前动物、植物及水产生物遗传育种领域的研究热点. 本文综述了基因组选择的计算方法、准确性的来源及影响因素以及基因组选择在畜禽遗传育种中的应用现状, 同时对当前存在的问题进行总结, 并对基因组选择的理论发展和应用前景进行展望.

关键词

基因组选择
基因组育种值
标记辅助选择
畜禽
育种

畜禽的多数重要经济性状都是数量性状. 传统的畜禽育种工作使用表型和系谱记录, 通过最佳线性无偏预测(best linear unbiased prediction, BLUP)来估计育种值(estimated breeding values, EBVs), 并在此基础上进行选择, 从而实现对畜禽重要经济性状的遗传改良. 此方法已经在畜禽遗传改良中取得了很大进展, 如通过使用传统育种手段, 美国奶牛的平均产奶性状在过去的 40 年间增加了 5000 kg, 其中有至少 60% 的贡献来自遗传改良. 但是, 对于一些遗传力低的性状的限性性状、长寿性状以及表型难以测定或测定成本较高的性状来说却进展缓慢. 近年来, 随着分子遗传学的发展, 遗传标记逐渐用于畜禽的标记辅助选择(marker assisted selection, MAS)育种中, 以提高遗传选育的准确性. 然而, 目前发现并经过功能确认的影响畜禽重要经济性状的基因或标记数量有限, 且它们解释遗传变异的比例较低. 这些都限制了标记辅助选择在畜禽育种中的应用^[1].

随着畜禽基因组研究的推进, 大多畜禽物种的

高密度全基因组 SNP (single nucleotide polymorphism, 单核苷酸多态)芯片已经面市. 为了将高密度全基因组标记信息更有效地用于遗传改良, Meuwissen 等人^[2]于 2001 年首次提出了基因组选择(genomic selection, GS)方法. 近年来, 针对基因组选择的计算方法、影响因素、应用策略和育种方案的研究大量涌现. 在第 9 届遗传学应用于畜牧生产世界大会(9th World Congress of Genetics Applied on Livestock Production, Leipzig, Germany)上, 有超过三分之一的报告与基因组选择有关^[3]. 这充分说明基因组选择方法已经成为世界畜禽^[1]、作物^[4]和水产生物^[5]遗传育种领域新的研究热点. 目前, 中国也已经启动相关的科研项目对基因组选择的方法及其在牛、猪、鸡及其他畜种中的应用进行研究. 针对基因组选择在畜禽遗传育种领域的研究进行综述尤为重要. 本文综述了基因组选择的计算方法、基因组育种值准确性的影响因素以及基因组选择在畜禽遗传育种中的应用现状, 同时对基因组选择中存在的问题进行探讨, 并对基因组选择的发展进行展望.

1 基因组选择的基本思想

育种值估计是动物遗传育种的核心内容之一。育种值估计方法的实质就是利用个体本身和(或)亲属的性状记录,进行适当加权来提高选择的准确性。标记辅助选择主要是将影响目标性状的基因或标记信息加入到遗传评估中,来提高育种值估计的准确性。然而,标记信息所能带来的额外准确性主要取决于它能够解释的遗传变异。畜禽遗传改良的多数目标性状都是数量性状,受多个基因控制,每个基因只能解释很小比例的遗传变异。因此,通过候选基因(candidate gene)、数量性状基因座定位(quantitative trait loci mapping, QTL mapping)和全基因组关联分析(genome-wide association study, GWAS)等策略发现的基因或标记也只能解释较小比例的遗传变异。显然,通过此策略实施标记辅助选择难以显著提高育种值估计的准确性。

基因组选择方法的提出则解决了标记辅助选择所面临的上述问题。从传统选择方法到基因组选择的策略变化如图1所示。基因组选择也是一种标记辅助选择,但与常规的标记辅助选择中只使用少数标记不同的是,基因组选择同时使用覆盖全基因组的标记进行育种值估计^[6],由此得到的估计育种值称为基因组育种值(genomic estimated breeding value, GEBV)。基因组选择的一个基本假设是,影响数量性状的每一个QTL都与高密度全基因组标记图谱中的至少一个标记处于连锁不平衡(linkage disequilibrium, LD)状态。因此,基因组选择能够追溯到所有影响目标性状的QTL,从而克服传统标记辅助选择中标记解释遗传方差较少的缺点,实现对育种值的准确预测。

2 基因组育种值的计算方法

按照所使用统计模型的不同,目前基因组育种

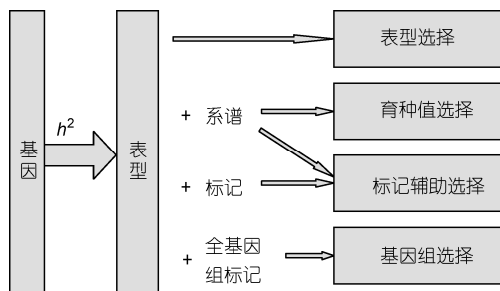


图1 动物育种选择方法示意图

值的计算方法主要可分为2类:第一类方法通过估计标记效应进而间接获得基因组估计育种值;第二类方法用标记构建个体间关系矩阵,将关系矩阵放入混合模型方程组(mixed model equations, MME),直接获得个体的基因组估计育种值。两类方法的基本计算模式如图2所示。基于这些方法,为满足实际数据处理的需求,也有研究对模型进行了扩展。

2.1 间接估计基因组育种值

间接估计基因组育种值的方法分为两步:(1)构建一定规模的参考群(reference population),利用参考群个体的表型和全基因组标记基因型信息估计全基因组中每一个标记对选择性状的效应值;(2)检测候选个体的全基因组标记基因型,依据其每个位点的标记基因型将由参考群体估计的标记效应累加获得个体的基因组育种值。标记效应估计模型为

$$y = Xb + \sum_{i=1}^m Z_i g_i + e, \quad (1)$$

其中, y 是参考群中所有个体的表型值向量; b 是固定效应向量; g_i 是第 i 个标记的随机效应; m 是总的标记数; X 和 Z 是关联矩阵; e 是随机残差向量,其方差-协方差矩阵为 $\sigma_e^2 I$, σ_e^2 是残差方差。

基于模型(1)的标记效应估计方法主要有:岭回归最佳线性无偏预测法^[2](ridge regression best linear unbiased prediction, RRBLUP)、贝叶斯方法A和B^[2](BayesA, BayesB)以及贝叶斯压缩^[7](Bayesian shrinkage, BayesS)。这些方法的差别主要在于对标记效应方差 σ_g^2 的分布的假设不同。其中RRBLUP假定所有标记都有效应,且方差相同;BayesA假定所有标记都有效应,且效应方差服从逆卡方分布(inversed chi-square distribution);BayesB则引入一个新参数 π ,

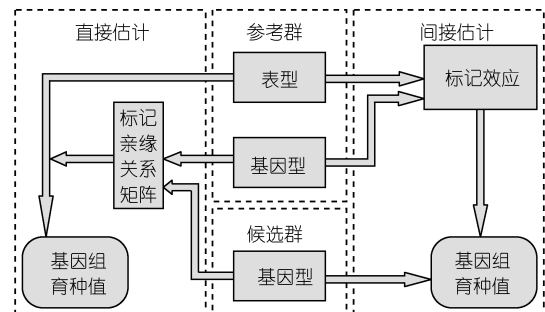


图2 基因组选择中直接和间接估计基因组育种值计算模式示意图

假定标记效应方差以 π 的概率值为0,以 $(1-\pi)$ 的概率服从逆卡方分布^[2]. 贝叶斯压缩对于标记效应方差的分布假设与 BayesA 相同,但是在效应估计时使用压缩算法^[7]. 目前的模拟研究结果表明, BayesB 方法在多数情况下要优于其他方法^[2]. 这可能是由于模拟数据中 QTL 数量有限,与 BayesB 方法的理论假设比较吻合^[8].

为避免 BayesB 方法中人为设定参数 π 对效应估计产生影响,又有研究者对此方法进行改进,称为 BayesC 方法^[9,10]. Verbyla 等人^[9]应用随机搜索变量选择(stochastic search variable selection, SSVS)方法在估计标记效应的同时,将 π 作为模型中的变量进行求解; Meuwissen^[10]则将标记效应假定为2个方差不等的正态分布的混合分布,把方差参数和 π 作为模型变量求解. 同时,这些改进也对 BayesB 的计算效率有所提高.

BayesB 方法使用高计算强度的 Metropolis-Hasting 和 Gibbs 抽样算法,因此计算时间较长. 为提高计算效率, Meuwissen 等人^[11]提出基于条件期望迭代(iteration conditional expectation)算法的 fBayesB (fast BayesB)方法. fBayesB 方法准确性虽略低于 BayesB,但却能大大地缩短计算时间.

除上述方法以外,还有研究通过机器学习(machine learning)^[12]、主成分分析(principle component analysis)^[13]及最小二乘回归(least square regression)^[2]等降维的方式减少模型中的变量数,进而估计基因组育种值. 还有研究使用半参数^[14](semiparametric)、非参数^[15](nonparametric)及贝叶斯 LASSO (least absolute shrinkage and selection operator)^[16]方法. 针对 RRBLUP 方法, VanRaden^[17]提出了非线性回归的方法,调整每个标记效应的方差,从而在 BLUP 计算中实现对标记效应的压缩求解.

在获得标记效应估计值(\hat{g}_i)后,候选群中第 j 个个体的基因组育种值则可由 $\sum_{i=1}^m Z_{ji} \hat{g}_i$ 算得.

2.2 直接估计基因组育种值

BLUP 方法是当前动物遗传评估的常规方法,在各国不同畜禽品种的遗传评估中广泛应用. 其优势之一在于能够充分利用所有动物个体的信息,实现这一点的关键是通过系谱构建分子亲缘关系矩阵(numerator relationship matrix, NRM 或 A 矩阵),以

此来反映个体间的遗传关系,但这种由系谱计算的遗传联系只是期望的遗传关系,而实际的遗传关系可能会由于孟德尔抽样离差而偏离期望值. 与系谱相比,利用分子标记可以更真实地描述个体间的遗传关系^[18]. 因此在全基因组标记可用的情况下,也可以用标记构建个体间的关系矩阵(G 矩阵)来取代 A 矩阵,实现基因组育种值的直接估计. 其模型为

$$y = Xb + Zu + e, \quad (2)$$

其中 u 是基因组育种值向量,其方差协方差矩阵为 $\sigma_a^2 G$, σ_a^2 是加性遗传方差, G 是根据标记构建的个体间的关系矩阵.

VanRaden 等人^[19]和 Habier 等人^[20]分别阐述了基于模型(2)的基因组选择方法的基本原理. 由于此方法将标记构建的 G 矩阵用于 BLUP 计算,为区别于传统 BLUP 和基于模型(1)的 RRBLUP,它被称为 GBLUP (genomic best linear unbiased prediction)方法. GBLUP 与基于模型(1)的方法相比具有以下几个优势:有效降低估计方程组的个数,降低计算强度;有表型及无表型的个体的育种值可以放在同一个模型中,以传统动物模型 BLUP 的方式直接估计;可以计算个体基因组育种值的可靠性(reliability). 基于模型(2)的基因组育种值可靠性的计算方法与传统 BLUP 中育种值可靠性的方法相同^[17,21]. GBLUP 方法的重点在于如何用全基因组标记来构建 G 矩阵. VanRaden^[17]提出多种方法构建 G 矩阵,并对这些方法进行模拟研究. Goddard^[22]和 Hayes 等人^[23]从理论上证明 GBLUP 方法等价于 RRBLUP 方法. 这主要是由于两方法使用相同的信息,并且都假定每个标记都有效且效应方差相等. 研究证明,GBLUP 要优于传统的利用表型和系谱的 BLUP 方法^[20,24,25],这是因为 GBLUP 中所用的 G 矩阵比传统的 BLUP 方法所用的 A 矩阵更真实地反映了个体间的遗传关系^[17,20,26].

考虑到影响遗传评估目标性状的所有基因的位置在全基因组中分布不均匀以及基因效应不等的情况, Zhang 等人^[27]提出,在用全基因组标记构建个体间关系矩阵时应对不同标记给予不同的权重. 标记的权重则来自于模型(1)中估计的标记效应. 由于标记效应是性状特异的,不同性状有不同的关系矩阵. 因此,以此方式构建的关系矩阵被称为性状特异的关系矩阵(trait-specific relationship matrix, TA 矩阵),应用 TA 矩阵进行基因组育种值估计的方法被称为 TABLUP. 模拟及真实数据分析都表明, TABLUP 的

预测能力高于 RRBLUP 和 GBLUP^[28], 这是因为 TA 矩阵考虑了遗传评估目标性状的遗传结构^[27].

2.3 模型的扩展

基因组选择假定高密度标记与影响目标性状的所有位点都处于较强的连锁不平衡. 所以, 当标记密度较低时基因组选择的应用效果会受到影响^[29]. 针对此种情况, 有研究提出了两种策略: 可在模型中加入多基因效应(polygenic effect)和估计单倍型(haplotype)效应.

将多基因效应加入模型(1)后, 其模型为

$$y = Xb + \sum_{i=1}^m Z_i g_i + Wu + e, \quad (3)$$

其中 u 是动物个体的剩余多基因效应, W 是 u 的关联矩阵, 其方差-协方差矩阵为 $\sigma_u^2 A$, σ_u^2 为多基因效应方差. Calus 等人^[30]的模拟结果表明, 在标记密度较低时使用模型(3)可提高育种值估计的准确性. 对于模型(2), VanRaden^[17]则提出将 A 和 G 矩阵进行加权求和, $wG+(1-w)A$, 从而将多基因效应直接包括到模型中. 但是, 基因组选择使用的全基因组标记信息已经包含了平均亲缘信息^[20], 此种扩展能否提高基因组育种值估计的准确性还有待验证.

单倍型效应估计的模型与模型(1)相同, 区别在于 g_i 是第 i 个单倍型效应值, 而 Z 是从个体到每个区段的每种单倍型的关联矩阵. Calus 等人^[25]研究了不同方式定义单倍型对基因组选择的影响. 结果表明, 在低标记密度时, 使用较长区段构建单倍型可以提高基因组育种值估计的准确性. 然而, 在真实数据处理中, 估计单倍型效应前需要进行标记的连锁相推断, 并对每个个体进行单倍型推断. 这会增加计算难度, 并且单倍型推断错误可能会降低此方法的优势. Habier 等人^[31]则在模型中加入对单倍型区段两端父源和母源标记等位基因传递概率(probabilities of descent of marker alleles), 并对此方法进行模拟研究. 结果表明, 此方法准确性虽然不及高密度标记基因组选择, 但是却能降低标记测定成本.

对于实际应用中某些重要个体没有基因型记录的问题, Legarra 等人^[32]提出新的 G 矩阵构建策略, 将模型(2)的 G 矩阵扩展到没有基因型记录的个体, 新的矩阵被称为 H 矩阵. 2011年, Chen 等人^[33]用一个肉鸡群体的数据验证了此种方法的有效性.

3 基因组育种值的准确性

3.1 准确性的预测

基因组育种值的准确性是基因组选择研究中备受关注的问题. 它是指基因组育种值与真实育种值(true breeding values, TBV)的相关系数^[2]. 在模拟研究中, 真实育种值已知, 准确性可以直接获得. 但是, 在实际应用中, 真实育种值未知, 且应用条件纷繁复杂. 因此, 在一定条件下实现对基因组育种值的准确性进行预测, 对基因组选择的育种方案设计和应用至关重要.

Daetwyler 等人^[34]推导出利用全基因组标记估计基因组育种值的准确性(r)的理论公式:

$$r = \sqrt{N_p h^2 / (N_p h^2 + N_G)}, \quad (4)$$

其中 N_p 指参考群体的表型记录数, h^2 指性状的遗传力, N_G 指影响目标性状的独立 QTL 数.

Goddard^[22]推导出形式不同的另一个理论公式:

$$r = \sqrt{1 - \frac{\lambda}{2N_p \sqrt{a}} \log \left(\frac{1+a+2\sqrt{a}}{1+a-2\sqrt{a}} \right)}, \quad (5)$$

其中 $a = 1+2\lambda/N_p$, $\lambda = (1-h^2)M_e/(h^2 \log(2N_e))$, N_e 是有效群体大小, M_e 是有效染色体片段数(effective number of chromosome segments), 可由下式近似估计:

$$M_e = (2N_e L) / \log(4N_e L), \quad (6)$$

其中 L 是染色体长度(Morgan).

基于有效染色体片段的概念, Daetwyler 等人^[35]进一步改进了他们的预测公式. 他们认为 GBLUP 方法不受性状遗传结构的影响, 其准确性为

$$r = \sqrt{N_p h^2 / (N_p h^2 + M_e)}, \quad (7)$$

而 BayesB 则受性状 QTL 数(N_{qtl})的影响, 其准确性为

$$r = \sqrt{N_p h^2 / (N_p h^2 + \min(M_e, N_{qtl}))}. \quad (8)$$

模拟结果显示, 预测准确性要略高于实际的准确性, 但二者的变化规律一致^[35,36].

值得注意的是, 为便于理论推导, 研究者做了多个理论假设^[22,34]. 这些假设可能会使公式在实际应用中受到诸多限制. 然而, 这些研究为我们进行基因组选择的理论分析和实验设计提供了参考依据.

3.2 影响因素

虽然上述预测公式还不能实现对具体应用个案的基因组育种值准确性进行准确预测, 但是这些研

究揭示了影响基因组选择的关键因素. 根据这些因素在基因组选择育种方案中是否具有可操控性, 可以将其分为以下两类:

(i) 不可变因素. 不可变因素主要包括染色体的长度、有效群体大小和目标性状的遗传结构. 染色体长度因物种而异, 有效群体大小与基因组选择所针对的群体的发展历史及当前的群体结构有关, 目标性状的遗传结构主要由性状的遗传力和影响性状的 QTL 数量决定, 这些因素都是客观存在的, 不能人为改变.

由公式(6)可知, 染色体的长度(L)和有效群体大小(N_e)决定了有效染色体片段数 M_e , L 与 N_e 的乘积越大, M_e 就越大, 而 M_e 越大, 则基因组育种值的准确性就越低. 此外, 标记间的连锁不平衡程度也与有效群体大小有关, 若两个标记间的重组率为 d , 则它们之间的连锁不平衡(r^2)的期望值^[37]为

$$r^2 = 1/(1 + 4N_e d), \quad (9)$$

由此可知, 有效群体大小越大, 连锁不平衡水平就越低, 进而影响着基因组选择的准确性.

由公式(4)或(7)可知, 在 QTL 数或有效染色体片段数相同的情况下, 若要获得相同的准确性, 遗传力为 0.1 的性状需要的参考群体的表型记录数(N_p)是遗传力为 0.5 的性状所需表型记录数的 5 倍.

(ii) 可变因素. 可变因素是基因组育种值估计中需要考虑并能人为调控的因素. 这主要包括: 参考群的规模和结构、标记的数量和密度、连锁不平衡的变化以及所使用的计算方法等.

参考群的规模直接决定了表型记录数 N_p 的大小. N_p 的大小会对估计的准确性起到决定性的作用. 模拟研究的结果^[2]与公式(4)和(5)的预测趋势基本一致. 除参考群的规模以外, 其群体组成也会影响基因组选择的准确性. Muir^[24]的研究表明, 在 N_p 恒定时, 参考群个体分布于多个世代比只分布在一个世代的基因组育种值的准确性高.

基因组选择的优势来自于高密度的全基因组标记. 标记密度取决于标记数量. 由公式(9)可知, 标记密度越高, 标记间的连锁不平衡水平越高, 因此期望获得的准确性也越高. 然而, 关于标记密度对基因组选择影响的探讨已经很多^[10,25,29], 这里不再赘述.

虽然标记密度决定了标记的连锁不平衡水平和标记效应估计的准确性, 但是这并不意味着在候选群中使用相同的标记密度, 基因组育种值的准确性

就会恒定不变. 在多个世代中应用基因组选择时, 标记和 QTL 间连锁不平衡的变化会导致准确性的下降^[2,20]. 重组(recombination)、选择(selection)和迁移(migration)会是引起畜禽群体连锁不平衡变化的主要因素. 针对这些影响, 可以通过提高标记密度、不断增加新标记、增加有表型的新个体和重估标记效应来降低准确性的损失. 这种变化也会影响标记效应的重估频率.

已有研究者对各种计算方法在不同情况下的表现进行比较^[8,38]. 由于各种计算方法的假设不同, 其适用范围会有差异. 而真实的物种、群体及性状也千差万别, 因此在不同的情况下应用基因组选择, 需要对方法进行选择和验证, 从而最大限度地发挥基因组选择的优势.

4 基因组选择的应用现状

基因组选择可以实现早期选种、缩短世代间隔, 从而更大程度的提高畜禽的遗传进展, 为育种机构带来更多经济效益^[6]. 2006 年, 伴随牛全基因组序列的公布, Schaeffer^[39]对基因组选择在奶牛育种体系中的应用进行经济学分析和探讨. 结果显示, 基因组选择在奶牛育种体系上的应用将会降低 92% 的育种成本^[39], 这引起了育种研究者和育种企业的关注. 此后, 基因组选择的研究和应用报道不断涌现.

4.1 在奶牛中的应用

基因组选择已经在奶牛群体中大规模应用^[40,41]. 据 Interbull (<http://www.interbull.org/>)对其成员国的调查^[42]显示, 至 2010 年有 11 个成员国在其国家奶牛育种群中应用基因组选择. 同时, 更多国家正在计划实施基因组选择. 基因组选择应用的性状几乎包含了目前奶牛育种目标中的所有性状. 基因型测定范围包括从验证公牛到泌乳母牛和小母牛. 如果说这些行动反应了奶牛育种者对基因组选择应用效果的乐观判断, 那么不断增多的对应用结果的报道则验证了其判断的正确性. 目前, 文献报道基因组选择计算结果的奶牛群体主要来自于美国-加拿大^[41]、澳大利亚^[9,38,43]、挪威^[44]、新西兰^[45]、荷兰^[46]、丹麦^[47]、德国^[48]和爱尔兰^[49]. 随着奶牛育种工作的推进, 各国参考群规模也不断增大. 表 1 总结了部分国家基因组选择所使用的群体规模、标记数量及计算方法, 其中部分内容根据 Interbull 最新公布数据进行了更新.

表 1 文献报道的基因组选择在奶牛群体中应用状况统计

国家(公司)	参考群 ^{a)}	实际标记数(芯片)	计算方法
美国 ^[41]	18080(H), 2828(J), 1331(B)	43382 (Illumina 54 K)	GBLUP, Nonlinear-BLUP
加拿大 ^[41]	8800(H)	43382 (Illumina 54 K)	GBLUP, Nonlinear-BLUP
澳大利亚 ^[9]	1098(H)	39048 (Illumina 54 K)	BayesA, BayesB, SSVS, BLUP
澳大利亚 ^[38]	1239	7372 (Affy 25 K)	FR-LS, RRBLUP, BayesS, SVR, PLSR
澳大利亚 ^[43]	781(H), 287(J)	39048 (Illumina 54 K)	GBLUP, BayesA, SSVS
挪威 ^[44]	500(R)	18991 (Affymetrix 25 K)	RRBLUP, BayesB, BayesC
新西兰 ^[45]	2626(H), 1639(J), 642(HxJ)	44146 (Illumina 54 K)	GBLUP
荷兰 ^[46]	16173(H)	46529 (CRV Illumina 60 K)	Bayes
丹麦/瑞典/芬兰 ^[47]	10217(H)	38055 (Illumina 54 K)	SSVS, GBLUP
法国	20918(H), 6835(M), 4970(N)	(Illumina 54 K)	-
德国 ^[48]	17477(H)	45181 (Illumina 54 K)	GBLUP, BayesB, BLUP
爱尔兰 ^[49]	4300	42598 (Illumina 54 K)	GBLUP

a) 瑞士褐牛(B), 荷斯坦牛(H), 娟山牛(J), 蒙贝利亚牛(M), 诺曼底牛(N), 挪威红牛(R)

从这些应用结果来看, 基因组育种值的准确性高于传统育种值^[41,48]. 对于多数性状, 基因组育种值在使用全部标记的不同计算方法间没有明显的差异^[38,41]. 但是, 对于乳脂率性状, 假定不同标记间效应方差不等的方法准确性要高于使用均匀先验的方法(如 RRBLUP)^[9,41]. 这主要是因为乳脂率性状的遗传结构不同于其他性状. 位于奶牛 14 号染色体上的 DGAT1 基因影响乳脂率并解释较大比例的遗传方差, 使得标记效应的分布明显偏离微效多基因模型. 基于模型(2)的 GBLUP 方法在参考群规模较小时的表现优于 Bayes 方法^[43], 但是当参考群和验证群间的亲缘关系减弱时, GBLUP 方法的准确性下降速度比 Bayes 方法的快^[43,48]. 群体内的比较^[41,44]和不同群体间的比较都表明: 参考群的规模越大, 其结果的准确性也越高. 还有研究对标记数据进行筛选, 获得均匀分布的低密度标记, 结果显示标记密度的降低会导致准确性的下降^[44]. 但是, 标记密度增高对结果的影响要小于参考群规模增大对结果的影响^[41].

目前, 在奶牛群体中基因组育种值的计算是结合传统遗传评估进行的. 上述国家参考群都是由验证公牛来构建, 并且大多使用传统育种值作为表型值来估计基因组育种值(表 1). 此时表型性状的观测遗传力(h^2)与育种值的可靠性(r^2)相当. 而验证公牛育种值可靠性都大于 0.90, 且奶牛的有效群体大小要小于其他畜种, 这极大地提升了基因组选择在奶牛群体中的应用效果. 在中国, 奶牛基因组选择研究也已于 2008 年展开. 目前, 已经初步建立了由约 2100 头母牛组成的参考群和 87 头后裔测定公牛组成

的验证群. 初步分析结果表明, 产奶性状的基因组育种值准确性为 0.60~0.75 (未发表数据).

4.2 在其他畜禽物种中的应用

van der Werf^[50]使用选择指数理论对基因组选择在肉羊和细毛羊群体中的应用进行模拟研究, 结果表明, 与传统育种相比, 基因组选择的应用使肉羊和细毛羊经济选择指数分别提高 30%和 40%. Daetwyler 等人^[51]使用 7180 只羊构成的参考群对其产肉和产毛性状进行了基因组预测, 结果表明美利奴羊的产毛性状基因组育种值准确性为 0.15~0.79, 产肉性状的准确性为-0.07~0.57. Gonzalez-Recio 等人^[52]用 333 和 61 只后裔测定的公鸡分别作为参考群和验证群对肉鸡饲料转化率进行研究. 结果显示, 所使用的基因组选择方法准确性均高于传统 BLUP. 同时, 使用依据信息熵减标准从 3481 个标记中筛选的 400 个低密度标记估计基因组育种值, 其准确性高于使用全部标记. Wolc 等人^[53]对 2708 只蛋鸡进行基因组选择研究, 结果表明, 基因组选择在蛋鸡早期和晚期选择中能够分别提高 100%和 88%的准确性.

目前, 基因组选择在猪、羊及鸡等畜禽中的应用范围不及奶牛, 且所见报道不多. 一方面可能与其全基因组芯片开发较晚有关. 然而, 猪(62163 SNPs; Illumina PorcineSNP60)、羊(54241 SNPs; Illumina OvineSNP50)、马(54602 SNPs; Illumina EquineSNP50)及鸡(Illumina iSelect 18K Custom genotype)的全基因组芯片已经可用, 这给基因组选择在这些畜种中的应用提供了更加便利的条件. 已有研究者对基因组

选择在猪^[54,55]、羊^[50,56]、鸡^[53,57,58]等中的应用策略进行探讨。同时,传统选择手段对于低遗传力、胴体以及限性性状等的选择进展缓慢,而基因组选择为加速这些性状遗传改良提供了契机。另一方面可能是由于高密度芯片应用成本较高。由于使用低密度芯片进行基因组选择可降低育种成本,因此受到诸多关注。Zhang 等人^[36]模拟研究表明:使用总标记数 5% 的标记即可获得高密度芯片 95% 的准确性,使用性状筛选标记效果好于均匀分布的标记。其他研究者在鸡^[58]、牛^[59]的群体中详细探讨了使用低密度芯片的效果和可行性问题。

5 基因组选择对传统育种体系的影响

5.1 改变当前畜禽育种模式

传统的遗传评估需要由系谱来记录个体间的亲缘关系,而基因组选择是以全基因组标记所记录的遗传及进化史为纽带将不同个体的表型记录联系起来。这种信息利用方式的改变深刻地影响着畜禽育种体系。对于奶牛以外的其他畜种,在传统的以“核心群-扩繁群-生产群”为构架的“金字塔”式的育种体系中,往往由于生产群与核心群不在同一个生产体系中,且不同等级的群体间没有系谱记录而无法将生产群中的表型记录用到核心群的遗传评估中。例如:生产群中杂种猪的生长性状记录和屠宰场中的胴体性状记录无法用于核心群纯种猪的遗传评估。然而,基因组选择则可以通过遗传标记追溯不同个体间的遗传联系(图 2),为不同区域、不同种群,甚至不同品种间的联合遗传评估提供可能。全基因组标记在这种体系下的应用不仅可以实现产品追溯并免于记录系谱^[60],而且可以将生产群的表现和市场需求的变化快速地反馈到核心群的育种工作中。以奶牛育种为例,传统奶牛的后裔测定育种体系中,公牛需等数年经过验证后才能使用,而如果此公牛验证成绩不好不能被使用则浪费了大量的饲养成本和管理成本。基因组选择则可以根据基因组信息对初生公牛进行早期选择,优秀青年公牛的冻精则可直接销售,其遗传优势更快地向后代传递^[61],这大大缩短了世代间隔,降低了育种成本。

同样,基因组选择也可以和传统的育种值选择相结合。以奶牛为例,可以仅根据基因组育种值或结合系谱指数进行早期选种,数年后,中选公牛又可利

用其大量的后代记录估计其育种值,并可以按一定比例对基因组育种值(GEBV)、系谱指数和育种值(EBV)进行合并,制定一个新的育种值指数,进行再次选择,成绩高的个体可扩大其使用频度和范围,成绩不理想则减少在育种方面的使用甚至淘汰。

5.2 可实现多品种联合遗传评估

由于不受系谱的限制,基因组选择育种体系中参考群个体既可以包括单品种、单群体,又可以包括多品种、多群体。de Roos 等人^[62]模拟的参考群来自于分别有 6, 30 和 300 个世代的遗传隔离的 2 个群体;而 Toosi 等人^[63]和 Ibánz-Escriche 等人^[60]模拟的参考群则来自于纯系、二元、三元或四元杂交。他们的模拟研究表明:来自多个群体或品种的个体共同组成的参考群要优于单一群体来源组成的参考群。然而,多品种的参考群需要高密度的标记来保证标记和 QTL 之间的连锁相在不同群体间是一致的。同时,使用高密度标记时,不用在模型中考虑群体或品种特异的 SNP 对结果的影响^[60]。Kizilkaya 等人^[64]使用真实的牛全基因组芯片数据和模拟的表型信息对种间使用基因组选择进行探讨,而 Hayes 等人^[43]则直接使用娟山牛和荷斯坦牛的混合参考群进行研究,结果的趋势与模拟研究结果一致。本课题组通过高密度标记间连锁不平衡状态建立不同群体间的遗传联系,进而合并数据,发现比单独进行基因组育种值预测效果更好(未发表结果)。

这种评估方式的改变,对中国的畜禽遗传评估也具有重要的现实意义。以中国的种猪遗传评估为例:中国商品猪种有大白、长白和杜洛克几大品种,种群规模小、分散、系谱记录不完整,且种畜间交流很少,这使得种猪联合育种工作很难开展。而通过使用基因组标记,可以直接建立各场间个体的遗传联系,进行联合评估,进而扩大遗传交流,提高多品种的评估准确性,整体提高我国猪育种水平。

5.3 促进不同育种机构紧密合作

基因组选择还为不同育种机构间的密切合作提供了更好的条件。由于大的参考群可以提高基因组选择的效果,而且不受系谱限制,所以不同育种机构间的合作变得更加可行。目前,美国加拿大的北美联合评估^[41]以及欧洲七国的 EuroGenomics 合作项目^[65]已经显示了国际合作育种的优势。除此之外,由于

GBLUP 方法在基因组育种值计算中无需使用原始的基因组标记信息(只需用标记构建 G 矩阵), 所以此方法还可以为不同机构间个体基因型信息的独立性和保密性提供保障^[66]。同时, Interbull 正在构建新的技术平台, 在保证各个国家和公司的育种数据相对独立和保密的情况下整合数据, 让所有的参与者都从更大的数据中受益。这些新的动向也给我国动物育种提供了新的思路。

除此之外, 基因组选择还能降低近交。2007 年, Daetwyler 等人^[26]对基因组选择导致的群体近交进行研究, 结果表明: 与同胞选择和 BLUP 选择相比, 基因组选择在提高育种值估计准确性的同时可以有效降低近交增量。这主要是由于基因组选择考虑了孟德尔离差, 增大了同胞间的差异, 从而降低了同胞被同时选做种用的概率, 进而降低近交增量。2009 年, Goddard^[22]认为, 若要使基因组选择的长期选择反应最大化, 应在指数中加大低频率有利等位基因的权重。

6 存在的问题和展望

畜禽基因组选择领域已经有很多的研究和应用成果, 然而这依旧是一个新兴的研究领域, 在理论和应用方面面临着诸多挑战。

6.1 计算方法

SNP 芯片技术还在不断发展, 更高密度的 SNP 芯片将不断出现, 如 Illumina 公司已推出了牛 800K 的商用 SNP 芯片(含有 780000 个 SNP)。同时, 下一代测序技术已经在植物的育种研究中得以大规模应用^[67,68]。此技术检测速度快、成本低, 若能够在畜禽育种中广泛应用, 将为畜禽基因组选择提供更丰富的遗传标记信息。随着 SNP 数量的增加, SNP 效应估计的计算难度将大大增加。比如, 与 54K 芯片相比, 使用牛的 800K SNP 芯片, RRBLUP 方法运行时内存需求将会增加~200 倍, 而 Bayes 方法的计算时长也将增加相应倍数。

除此之外, 当前的基因组选择模型, 尤其是模型(1), 只能处理单性状、单次记录的数据。在处理一个个体有多性状记录或者一个性状有多次记录的数据时, 往往由于同一个体只有一组遗传标记, 导致系数矩阵会出现共线性问题而无法求解。但是, 传统的 BLUP 方法对于多性状、多信息来源及重复记录的处理

非常成熟, 且能够有效地提高育种值估计的准确性, 这也可能会成为基因组选择新方法的研究方向。

6.2 标记构建的关系矩阵

面对更高密度的 SNP 数据, 可以先使用标记构建个体间的关系矩阵, 从而将方程组的维数从标记数降低到个体数大小, 并用模型(2)直接求解基因组育种值。然而, 即使通过此方式降低了方程组求解的计算强度, 关系矩阵的构建和求逆也面临困难。传统 BLUP 使用系谱直接构建 A^{-1} 矩阵, 从而实现快速求解。而由标记构建的关系矩阵不是稀疏矩阵, 不仅需要进行高计算强度的标记关联矩阵相乘计算, 而且还要进行矩阵的求逆运算。这些问题有待使用并行计算或其他新的计算方法来解决。

6.3 其他问题

在应用方面, 基因组选择正逐渐进入奶牛遗传评估体系, 其在奶牛中的应用环境条件较为成熟。然而, 基因组选择在其他畜禽中的应用面临着诸多困难。其中, 最主要的一个问题是应用成本仍然较高。当前, 畜禽全基因组标记测定成本较高, 一个样本的全基因组芯片测定费用是 190~250 美元。与乳用种公牛相比, 其他种畜禽的种公畜影响力较小, 经济价值较低, 在猪、鸡育种中采用基因组选择方法的经济回报远没有奶牛中那么明显。虽然大批量使用高密度芯片测定基因型会增加育种企业的投入, 可能降低育种企业短期经济效益。但是, 新生物技术促使基因型检测费用的降低, 以及只在大型育种公司的核心群中使用基因组选择可能会为其带来可观的效益^[8]。在基因型检测费没有降到足够“低”之前, 低密度标记基因组选择方法^[31,69]的应用也可能会给其他畜禽育种带来新的契机。

同时, 在奶牛中可以使用高可靠性的估计育种值($r^2 > 0.90$)作为表型值估计基因组育种值, 但是这在其他畜禽中还不现实。因此, 与在奶牛中的应用相比, 在这些畜禽中的应用可能需要构建更大的参考群才能获得同奶牛中相同的准确性。即使不考虑测定成本, 由于缺乏高可靠性的育种值, 基因组选择方法也难以在这些群体中直接得到可靠的验证, 这无疑会增加基因组选择应用效果的不确定性。因此, 新的验证方法也有待进一步研究。

6.4 展望

虽然基因组选择的计算和应用仍面临诸多挑战,但毋庸置疑的是,基因组选择方法正在改变全世界奶牛遗传育种体系,并积极影响着其他畜禽的遗传育

种进程.它向人们证实了基因组时代分子生物学技术和计算技术已成为推动动物育种领域前进的强大动力.随着相关技术的不断进步,基因组选择方法会在动物育种领域得到更大范围的推广和应用,并必将给动物育种和遗传学研究领域带来深远的影响.

参考文献

- 1 Goddard M E, Hayes B J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet*, 2009, 10: 381–391
- 2 Meuwissen T H E, Hayes B J, Goddard M E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 2001, 157: 1819–1829
- 3 Habier D. More than a third of the WCGALP presentations on genomic selection. *J Anim Breed Genet*, 2010, 127: 336–337
- 4 Heffner E L, Sorrells M E, Jannink J L. Genomic selection for crop improvement. *Crop Sci*, 2009, 49: 1–12
- 5 Sonesson A K, Meuwissen T H. Testing strategies for genomic selection in aquaculture breeding programs. *Genet Sel Evol*, 2009, 41: 37
- 6 Goddard M E, Hayes B J. Genomic selection. *J Anim Breed Genet*, 2007, 124: 323–330
- 7 Xu S. Estimating polygenic effects using markers of the entire genome. *Genetics*, 2003, 163: 789–801
- 8 Calus M P L. Genomic breeding value prediction: Methods and procedures. *Animal*, 2010, 4: 157–164
- 9 Verbyla K L, Hayes B J, Bowman P J, et al. Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. *Genet Res*, 2009, 91: 307–311
- 10 Meuwissen T H. Accuracy of breeding values of ‘unrelated’ individuals predicted by dense SNP genotyping. *Genet Sel Evol*, 2009, 41: 35
- 11 Meuwissen T H, Solberg T R, Shepherd R, et al. A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. *Genet Sel Evol*, 2009, 41: 2
- 12 Long N, Gianola D, Rosa G J, et al. Machine learning classification procedure for selecting SNPs in genomic selection: Application to early mortality in broilers. *J Anim Breed Genet*, 2007, 124: 377–389
- 13 Solberg T R, Sonesson A K, Woolliams J A, et al. Reducing dimensionality for prediction of genome-wide breeding values. *Genet Sel Evol*, 2009, 41: 29
- 14 Gianola D, Fernando R L, Stella A. Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*, 2006, 173: 1761–1776
- 15 Bennewitz J, Solberg T, Meuwissen T. Genomic breeding value estimation using nonparametric additive regression models. *Genet Sel Evol*, 2009, 41: 20
- 16 de los Campos G, Naya H, Gianola D, et al. Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*, 2009, 182: 375–385
- 17 VanRaden P M. Efficient methods to compute genomic predictions. *J Dairy Sci*, 2008, 91: 4414–4423
- 18 Visscher P M, Medland S E, Ferreira M A, et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet*, 2006, 2: 316–325
- 19 VanRaden P M, Tooker M E. Methods to explain genomic estimates of breeding value. *J Dairy Sci*, 2007, 90(Suppl 1): 374
- 20 Habier D, Fernando R L, Dekkers J C M. The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 2007, 177: 2389–2397
- 21 Strandén I, Garrick D J. Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J Dairy Sci*, 2009, 92: 2971–2975
- 22 Goddard M E. Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica*, 2009, 136: 245–257
- 23 Hayes B J, Visscher P M, Goddard M E. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet Res*, 2009, 91: 47–60
- 24 Muir W M. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J Anim Breed Genet*, 2007, 124: 342–355
- 25 Calus M P, Meuwissen T H, de Roos A P, et al. Accuracy of genomic selection using different methods to define haplotypes. *Genetics*, 2008, 178: 553–561
- 26 Daetwyler H D, Villanueva B, Bijma P, et al. Inbreeding in genome-wide selection. *J Anim Breed Genet*, 2007, 124: 369–376

- 27 Zhang Z, Ding X D, Liu J F, et al. TA-BLUP: A new genetic evaluation method for genomic selection. In: 9th World Congress on Genetics Applied to Livestock Production, 2010. Leipzig, Germany. 215
- 28 Zhang Z, Liu J F, Ding X D, et al. Best linear unbiased prediction of genomic breeding values using trait-specific marker-derived relationship matrix. *PLoS ONE*, 2010, 5: e12648
- 29 Solberg T R, Sonesson A K, Woolliams J A, et al. Genomic selection using different marker types and densities. *J Anim Sci*, 2008, 86: 2447–2454
- 30 Calus M P L, Veerkamp R F. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. *J Anim Breed Genet*, 2007, 124: 362–368
- 31 Habier D, Fernando R L, Dekkers J C M. Genomic selection using low-density marker panels. *Genetics*, 2009, 182: 343–353
- 32 Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. *J Dairy Sci*, 2009, 92: 4656–4663
- 33 Chen C Y, Misztal I, Aguilar I, et al. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. *J Anim Sci*, 2011, 89: 23–28
- 34 Daetwyler H D, Villanueva B, Woolliams J A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE*, 2008, 3: e3395
- 35 Daetwyler H D, Pong-Wong R, Villanueva B, et al. The impact of genetic architecture on genome-wide evaluation methods. *Genetics*, 2010, 185: 1021–1031
- 36 Zhang Z, Ding X D, Liu J F, et al. Accuracy of genomic prediction using low density marker panels. *J Dairy Sci*, 2011, doi: 10.3168/jds.2010-3917
- 37 Sved J A. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor Popul Biol*, 1971, 2: 125–141
- 38 Moser G, Tier B, Crump R E, et al. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. *Genet Sel Evol*, 2009, 41: 56
- 39 Schaeffer L R. Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed Genet*, 2006, 123: 218–223
- 40 Hayes B J, Bowman P J, Chamberlain A J, et al. Genomic selection in dairy cattle: Progress and challenges. *J Dairy Sci*, 2009, 92: 433–443
- 41 VanRaden P M, Van Tassell C P, Wiggans G R, et al. Reliability of genomic predictions for North American Holstein bulls. *J Dairy Sci*, 2009, 92: 16–24
- 42 Loberg A, Dürr J W. Interbull survey on the use of genomic information. *Interbull Bull*, 2009, 39: 3–14
- 43 Hayes B, Bowman P, Chamberlain A, et al. Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genet Sel Evol*, 2009, 41: 51
- 44 Luan T, Woolliams J A, Lien S, et al. The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. *Genetics*, 2009, 183: 1119–1126
- 45 Harris B L, Montgomerie W A. Current status of the use of genomic information in the national genetic evaluation in New Zealand. In: Proc of the Interbull International Workshop in Genomic Information in Genetic Evaluations. Interbull: Uppsala, Sweden, 2009. 35–38
- 46 de Roos A P W, Schrooten C, Mullaart E, et al. Genomic selection at CRV. *Interbull Bull*, 2009, 39: 47–50
- 47 Lund M S, Su G. Genomic selection in the Nordic countries. *Interbull Bull*, 2009, 39: 39–42
- 48 Habier D, Tetens J, Seefried F R, et al. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet Sel Evol*, 2010, 42: 5
- 49 Berry D, Kearney F, Harris B. Genomic selection in Ireland. *Interbull Bull*, 2009, 39: 29–34
- 50 van der Werf J H J. Potential benefit of genomic selection in sheep. In: The Association for the Advancement of Animal Breeding and Genetics 18th Conference, 2009. 38–41
- 51 Daetwyler H D, Hickey J M, Henshall J M, et al. Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. *Anim Prod Sci*, 2010, 50: 1004–1010
- 52 Gonzalez-Recio O, Gianola D, Rosa G J, et al. Genome-assisted prediction of a quantitative trait measured in parents and progeny: Application to food conversion rate in chickens. *Genet Sel Evol*, 2009, 41: 3
- 53 Wolc A, Stricker C, Arango J, et al. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genet Sel Evol*, 2011, 43: 5
- 54 Tuggle C K, Dekkers J C M. Genotyping: How useful is it for producers? *Pig Prog*, 2009, 25: 9
- 55 Simianer H. The potential of genomic selection to improve litter size in pig breeding programmes. In: European Association for Animal Production, 2009
- 56 McEwan J C. Current and future impact of DNA technologies on the New Zealand sheep industry. In: Proceedings of the New Zealand Society of Animal Production, 2009. 165–168

- 57 Chen C Y, Misztal I, Aguilar I, et al. Genome wide marker assisted selection in chicken: Making the most of all data, pedigree, phenotypic, and genomic in a simple one step procedure. In: 9th World Congress on Genetics Applied to Livestock Production, 2010. Leipzig, Germany
- 58 Legarra A, Calenge F, Mariani P, et al. Use of a reduced set of single nucleotide polymorphisms for genetic evaluation of resistance to Salmonella carrier state in laying hens. *Poult Sci*, 2011, 90: 731–736
- 59 Weigel K A, de los Campos G, Vazquez A I, et al. Accuracy of direct genomic values derived from imputed single nucleotide polymorphism genotypes in Jersey cattle. *J Dairy Sci*, 2010, 93: 5423–5435
- 60 Ibáñez-Escriche N, Fernando R L, Toosi A, et al. Genomic selection of purebreds for crossbred performance. *Genet Sel Evol*, 2009, 41: 12
- 61 König S, Simianer H, Willam A. Economic evaluation of genomic breeding programs. *J Dairy Sci*, 2009, 92: 382–391
- 62 de Roos A P, Hayes B J, Goddard M E. Reliability of genomic predictions across multiple populations. *Genetics*, 2009, 183: 1545–1553
- 63 Toosi A, Fernando R L, Dekkers J C. Genomic selection in admixed and crossbred populations. *J Anim Sci*, 2010, 88: 32–46
- 64 Kizilkaya K, Fernando R L, Garrick D J. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *J Anim Sci*, 2010, 88: 544–551
- 65 Lund M S, Roos A P W d, Vries A G d, et al. Improving genomic prediction by EuroGenomics collaboration. In: 9th World Conference of Genetics Applied on Livestock Production, 2010. 880
- 66 Harris B L, Johnson D L, Montgomerie W A. National genomic evaluations without genotypes. *Interbull Bull*, 2009, 40: 189–192
- 67 Tian F, Bradbury P J, Brown P J, et al. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet*, 2011, 43: 159–162
- 68 Huang X H, Wei X H, Sang T, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet*, 2010, 42: 961–976
- 69 Weigel K A, de los Campos G, Gonzalez-Recio O, et al. Predictive ability of direct genomic values for lifetime net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers. *J Dairy Sci*, 2009, 92: 5248–5257